

Maximizing the Weighted Number of Spanning Trees: Near- t -Optimal Graphs*

Kasra Khosoussi[†] Gaurav S. Sukhatme[‡] Shoudong Huang[†] Gamini Dissanayake[†]

April 13, 2016

Abstract

Designing well-connected graphs is a fundamental problem that frequently arises in various contexts across science and engineering. The weighted number of spanning trees, as a connectivity measure, emerges in numerous problems and plays a key role in, e.g., network reliability under random edge failure, estimation over networks and D-optimal experimental designs. This paper tackles the open problem of designing graphs with the maximum weighted number of spanning trees under various constraints. We reveal several new structures, such as the log-submodularity of the weighted number of spanning trees in connected graphs. We then exploit these structures and design a pair of efficient approximation algorithms with performance guarantees and near-optimality certificates. Our results can be readily applied to a wide verity of applications involving graph synthesis and graph sparsification scenarios.

*Working paper. kasra.mail@gmail.com – <https://kasra.github.io>

[†]Centre for Autonomous Systems (CAS), University of Technology Sydney.

[‡]Department of Computer Science, University of Southern California.

Contents

1	Introduction	3
2	Background	3
2.1	Preliminaries	3
2.2	Matrix-Tree Theorems	4
3	Tree-Connectivity	5
4	ESP: Edge Selection Problem	7
4.1	Problem Definition	7
4.2	Exhaustive Search	8
4.3	Greedy Algorithm	8
4.4	Convex Relaxation	10
4.5	Certifying Near-Optimality	12
4.6	Numerical Results	13
5	Beyond k-ESP⁺	13
5.1	Matroid Constraints	13
5.1.1	Greedy Algorithm	14
5.1.2	Convex Relaxation	14
5.2	Dual of k -ESP ⁺	15
5.2.1	Greedy Algorithm	15
5.2.2	Convex Relaxation	16
5.2.3	Certifying Near-Optimality	17
6	Conclusion	17
A	Proofs	20

1 Introduction

Various graph connectivity measures have been studied and used in different contexts. Among them are the combinatorial measures, such as vertex/edge-connectivity, as well as spectral notions, like algebraic connectivity [10]. As a connectivity measure, the number of spanning trees (sometimes referred to as *graph complexity* or *tree-connectivity*) stands out in this list since despite its combinatorial origin, it can also be characterized solely based on the spectrum of graph Laplacian. It has been shown that tree-connectivity is associated with D-optimal (determinant-optimal) experimental designs [8, 6, 1, 26]. The number of spanning trees also appears in the study of *all-terminal network reliability* under (i.i.d.) random edge failure (defined as the probability of network being connected) [16, 30]. In particular, it has been proved that for a given number of edges and vertices, the *uniformly-most reliable* network, *upon existence*, must have the maximum number of spanning trees [3, 22, 4]. The graph with the maximum number of spanning trees among a finite set of graphs (e.g., graphs with n vertices and m edges) is called *t-optimal*. The problem of identifying *t-optimal* graphs under a (n, m) constraint remains open and has been solved only for specific pairs of (n, m) ; see, e.g., [27, 6, 15, 25]. We prove that the (weighted) number of spanning trees in connected graphs can be posed as a monotone log-submodular function. This structure enables us to design a complementary greedy-convex pair of approximate algorithms to synthesize near-*t-optimal* graphs under several constraints with approximation guarantees and near-optimality certificates.

Notation

Throughout this paper, bold lower-case and upper-case letters are reserved for real vectors and matrices, respectively. The standard basis for \mathbb{R}^n is denoted by $\{\mathbf{e}_i^n\}_{i=1}^n$, and \mathbf{e}_0^n is defined to be the zero n -vector. For any $n \in \mathbb{N}$, $[n]$ denotes the set $\mathbb{N}_{\leq n} = \{1, 2, \dots, n\}$. Sets are shown by upper-case letters. $|\mathcal{X}|$ denotes the cardinality of set \mathcal{X} . For any finite set \mathcal{W} , $\binom{\mathcal{W}}{k}$ is the set of all k -subsets of \mathcal{W} . The eigenvalues of symmetric matrix \mathbf{M} are denoted by $\lambda_1(\mathbf{M}) \leq \dots \leq \lambda_n(\mathbf{M})$. $\mathbf{1}$, \mathbf{I} and $\mathbf{0}$ denote the vector of all ones, identity and zero matrix with appropriate sizes, respectively. $\mathbf{S}_1 \succ \mathbf{S}_2$ means $\mathbf{S}_1 - \mathbf{S}_2$ is positive-definite. The Euclidean norm is denoted by $\|\cdot\|$. $\text{diag}(\mathbf{W}_i)_{i=1}^k$ is the block-diagonal matrix with matrices $(\mathbf{W}_i)_{i=1}^k$ as blocks on its main diagonal. For any graph G , $E(G)$ denotes the edge set of G . Finally, $\mathbb{S}_{\geq 0}^n$ and $\mathbb{S}_{> 0}^n$ denote the set of symmetric positive semidefinite and symmetric positive definite matrices in $\mathbb{R}^{n \times n}$, respectively.

2 Background

2.1 Preliminaries

Let $G = (V, E)$ be an undirected graph over $V = [n]$ and with $|E| = m$ edges. By assigning a positive weight to each edge of the graph through $w : E \rightarrow \mathbb{R}_{> 0}$, we obtain $G^w = (V, E, w)$. To shorten our notation let us define $w_{uv} \triangleq w(u, v) = w(v, u)$. As it will become clear shortly, without loss of generality we can assume G is a simple graph since (i) loops do not affect the number of spanning trees, and (ii) parallel edges can be replaced by a single edge whose weight is the sum of the weights of the parallel edges. $\mathbf{W} \triangleq \text{diag}(w(e_1), \dots, w(e_m))$ denotes the weight matrix in which $e_i \in E$ is the i th edge. The degree of vertex $v \in V$ in G is denoted by $\deg(v)$. Let $\tilde{\mathbf{A}}$ be the incidence matrix of G after assigning arbitrary orientations

to its edges. The Laplacian matrix of G is defined as $\tilde{\mathbf{L}} \triangleq \tilde{\mathbf{A}}\tilde{\mathbf{A}}^\top$. For an arbitrary choice of $v_0 \in V$, let $\mathbf{A} \in \{-1, 0, 1\}^{(n-1) \times m}$ be the matrix obtained by removing the row that corresponds to v_0 from $\tilde{\mathbf{A}}$. We call \mathbf{A} the *reduced incidence matrix* of G after *anchoring* v_0 . The *reduced Laplacian matrix* of G is defined as $\mathbf{L} \triangleq \mathbf{A}\mathbf{A}^\top$. \mathbf{L} is also known as the *Dirichlet* or *grounded Laplacian matrix* of G . Note that \mathbf{L} can also be obtained by removing the row and column associated to the anchor from the graph Laplacian matrix. \mathbf{A} is full column rank and consequently \mathbf{L} is positive definite, iff G is connected. For weighted graphs, $\mathbf{A}\mathbf{W}\mathbf{A}^\top$ is the *reduced weighted Laplacian* of G^w . Note that this is a natural generalization of \mathbf{L} , and will reduce to its unweighted counterpart if all weights are equal to one (i.e., $\mathbf{W} = \mathbf{I}$). The reduced (weighted) Laplacian matrix can be decomposed into the (weighted) sum of *elementary reduced Laplacian matrices*:

$$\mathbf{L} = \sum_{\{u,v\} \in E} w(u,v) \mathbf{L}_{uv} \quad (1)$$

in which $\mathbf{L}_{uv} \triangleq \mathbf{a}_{uv}\mathbf{a}_{uv}^\top$ and $\mathbf{a}_{uv} = \mathbf{e}_u - \mathbf{e}_v$ is the corresponding column of \mathbf{A} .

2.2 Matrix-Tree Theorems

The spanning trees of G are spanning subgraphs of G that are also trees. Let \mathcal{T}_G and $t(G) \triangleq |\mathcal{T}_G|$ denote the set of all spanning trees of G and its number of spanning trees, respectively. Let T_n and K_n be, respectively, an arbitrary tree and the complete graph with n vertices. The following statements hold.

1. $t(G) \geq 0$, and $t(G) = 0$ iff G is disconnected,
2. $t(T_n) = 1$,
3. $t(K_n) = n^{n-2}$ (Cayley's formula),
4. if G is connected, then $t(T_n) \leq t(G) \leq t(K_n)$,
5. if G_1 is a spanning subgraph of G_2 , then $t(G_1) \leq t(G_2)$.

Therefore $t(G)$ is a sensible measure of graph connectivity. The following theorem by Kirchhoff provides an expression for computing $t(G)$.

Theorem 2.1 (Matrix-Tree Theorem [10]). *Let \mathbf{L}_G and $\tilde{\mathbf{L}}_G$ be, respectively, the reduced Laplacian and the Laplacian matrix of any simple undirected graph G after anchoring an arbitrary vertex out of its n vertices. The following statements hold.*

1. $t(G) = \det(\mathbf{L}_G)$,
2. $t(G) = \frac{1}{n} \prod_{i=2}^n \lambda_i(\tilde{\mathbf{L}}_G)$.¹

The matrix-tree theorem can be naturally generalized to weighted graphs, where each spanning tree is “counted” according to its *value* defined below.

¹Recall that the Laplacian matrix of any connected graph has a zero eigenvalue with multiplicity one (see, e.g., [10]).

Definition 2.1. Suppose $G = (V, E, w)$ is a weighted graph with a non-negative weight function. The value of each spanning tree of G is measured by the following function,

$$\mathbb{V}_w : \mathcal{T}_G \rightarrow \mathbb{R}_{\geq 0} \quad (2)$$

$$T \mapsto \prod_{e \in E(T)} w(e). \quad (3)$$

Furthermore, we define the weighted number of trees as $t_w(G) \triangleq \sum_{T \in \mathcal{T}_G} \mathbb{V}_w(T)$.

Theorem 2.2 (Weighted Matrix-Tree Theorem [20]). *For every simple weighted graph $G = (V, E, w)$ with $w : E \rightarrow \mathbb{R}_{>0}$ we have $t_w(G) = \det \mathbf{A} \mathbf{W} \mathbf{A}^\top$.*

Note that Theorem 2.2 reduces to Theorem 2.1 if $w(e) = 1$ for all $e \in E$. Therefore, in the rest of this paper we focus our attention mainly on weighted graphs.

Definition 2.2. The weighted *tree-connectivity* of graph G is formally defined as

$$\tau_w(G) \triangleq \begin{cases} \log t_w(G) & \text{if } t_w(G) > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

3 Tree-Connectivity

Definition 3.1. Consider an arbitrary simple undirected graph G° . Let p_i be the probability assigned to the i th edge, and \mathbf{p} be the stacked vector of probabilities. $G \sim \mathbb{G}(G^\circ, \mathbf{p})$ indicates that

1. G is a spanning subgraph of G° .
2. The i th edge of G° appears in G with probability p_i , independent of other edges.

The naive procedure for computing the expected weighted number of spanning trees in such random graphs involves a summation over exponentially many terms. Theorem 3.1 offers an efficient and intuitive way of computing this expectation in terms of G° and \mathbf{p} .

Theorem 3.1. *For any $\mathbb{G}(G^\circ, \mathbf{p})$ and $w : E(K_n) \rightarrow \mathbb{R}_{>0}$,*

$$\mathbb{E}_{G \sim \mathbb{G}(G^\circ, \mathbf{p})} [t_w(G)] = t_{w_p}(G^\circ), \quad (5)$$

where $w_p(e_i) \triangleq p_i w(e_i)$ for all $e_i \in E(G^\circ)$.

Note that this expectation can now be computed in $\mathcal{O}(n^3)$ time for general G° .

Lemma 3.1. *Let G^+ be the graph obtained by adding $\{u, v\} \notin E$ with weight w_{uv} to $G = (V, E, w)$. Let \mathbf{L}_G be the reduced Laplacian matrix and \mathbf{a}_{uv} be the corresponding column of the reduced incidence matrix of G after anchoring an arbitrary vertex. If G is connected,*

$$\tau_w(G^+) = \tau_w(G) + \log(1 + w_{uv} \Delta_{uv}^G), \quad (6)$$

where $\Delta_{uv}^G \triangleq \mathbf{a}_{uv}^\top \mathbf{L}_G^{-1} \mathbf{a}_{uv}$.

Lemma 3.2. *Similar to Lemma 3.1, let G^- be the graph obtained by removing $\{p, q\} \in E$ with weight w_{pq} from E . If G is connected,*

$$\tau_w(G^-) = \tau_w(G) + \log(1 - w_{pq}\Delta_{pq}^G). \quad (7)$$

Corollary 3.2. *Define $\mathcal{T}_G^{uv} \triangleq \{T \in \mathcal{T}_G : \{u, v\} \in E(T)\}$. Then we have*

$$\Delta_{uv}^G = |\mathcal{T}_G^{uv}|/|\mathcal{T}_G| = |\mathcal{T}_G^{uv}|/t(G). \quad (8)$$

Similarly, for weighted graphs we have

$$w_{uv}\Delta_{uv}^G = \frac{\sum_{T \in \mathcal{T}_G^{uv}} \mathbb{V}_w(T)}{\sum_{T \in \mathcal{T}_G} \mathbb{V}_w(T)} = \frac{\sum_{T \in \mathcal{T}_G^{uv}} \mathbb{V}_w(T)}{t_w(G)}. \quad (9)$$

Lemmas 3.1 and 3.2 imply that $w_{uv}\Delta_{uv}^G$ determines the change in tree-connectivity after adding or removing an edge. This term is known as the *effective resistance* between u and v . If G is an electrical circuit where each edge represents a resistor with a conductance equal to its weight, then $w_{uv}\Delta_{uv}^G$ is equal to the electrical resistance across u and v . The effective resistance also emerges as a key factor in various other contexts; see, e.g., [9, 2, 19]. Note that although we derived Δ_{uv}^G using the *reduced* graph Laplacian, it is more common to define the effective resistance using the pseudoinverse of graph Laplacian $\tilde{\mathbf{L}}_G$ [9].

Now, on a seemingly unrelated note, we turn our attention to structures associated to tree-connectivity when seen as a set function.

Definition 3.2. Let V be a set of $n \geq 2$ vertices. Denote by G_E the graph (V, E) for any $E \in E(K_n)$. For any $w : 2^{E(K_n)} \rightarrow \mathbb{R}_{>0}$ define

$$\begin{aligned} \text{tree}_{n,w} : 2^{E(K_n)} &\rightarrow \mathbb{R}_{\geq 0} \\ E &\mapsto t_w(G_E), \end{aligned} \quad (10)$$

$$\begin{aligned} \log \text{tree}_{n,w} : 2^{E(K_n)} &\rightarrow \mathbb{R} \\ E &\mapsto \tau_w(G_E). \end{aligned} \quad (11)$$

Definition 3.3 (Tree-Connectivity Gain). Suppose a *connected* base graph (V, E_{init}) with $n \geq 2$ vertices and an arbitrary positive weight function $w : E(K_n) \rightarrow \mathbb{R}_{>0}$ are given. Define

$$\begin{aligned} \log \text{TG}_{n,w} : 2^{E(K_n)} &\rightarrow \mathbb{R}_{\geq 0} \\ E &\mapsto \log \text{tree}_{n,w}(E \cup E_{\text{init}}) - \log \text{tree}_{n,w}(E_{\text{init}}). \end{aligned} \quad (12)$$

Definition 3.4. Suppose \mathcal{W} is a finite set. For any $\xi : 2^{\mathcal{W}} \rightarrow \mathbb{R}$,

1. ξ is called *normalized* iff $\xi(\emptyset) = 0$.
2. ξ is called *monotone* if $\xi(\mathcal{B}) \geq \xi(\mathcal{A})$ for every \mathcal{A} and \mathcal{B} s.t. $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{W}$.

3. ξ is called *submodular* iff for every \mathcal{A} and \mathcal{B} s.t. $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{W}$ and $\forall s \in \mathcal{W} \setminus \mathcal{B}$ we have,

$$\xi(\mathcal{A} \cup \{s\}) - \xi(\mathcal{A}) \geq \xi(\mathcal{B} \cup \{s\}) - \xi(\mathcal{B}). \quad (13)$$

4. ξ is called *supermodular* iff $-\xi$ is submodular.

5. ξ is called *log-submodular* iff ξ is positive and $\log \xi$ is submodular.

Theorem 3.3. $\text{tree}_{n,w}$ is normalized, monotone and supermodular.

Theorem 3.4. $\log \text{TG}_{n,w}$ is normalized, monotone and submodular.

Corollary 3.5 follows directly from Theorems 3.1, 3.3 and 3.4.

Corollary 3.5. The expected weighted number of spanning trees in random graphs is normalized, monotone and supermodular when seen as a set function similar to $\text{tree}_{n,w}$. Moreover, the expected weighted number of spanning trees can be posed as a log-submodular function similar to $\log \text{TG}_{n,w}$.

4 ESP: Edge Selection Problem

4.1 Problem Definition

Suppose a connected base graph is given. The edge selection problem (ESP) is a combinatorial optimization problem whose goal is to pick the optimal k -set of edges from a given candidate set of new edges such that the weighted number of spanning trees after adding those edges to the base graph is maximized.

Problem 4.1 (ESP). Let $G_{\text{init}} = (V, E_{\text{init}}, w)$ be a given connected graph where $w : E(K_n) \rightarrow \mathbb{R}_{>0}$. Consider the following scenarios.

1. $k\text{-ESP}^+$: For some $\mathcal{M}^+ \subseteq E(K_n) \setminus E_{\text{init}}$,

$$\begin{aligned} & \underset{E \subseteq \mathcal{M}^+}{\text{maximize}} && t_w(G_{E_{\text{init}} \cup E}) \\ & \text{subject to} && |E| = k. \end{aligned} \quad (14)$$

2. $k\text{-ESP}^-$: For some $\mathcal{M}^- \subseteq E_{\text{init}}$,

$$\begin{aligned} & \underset{E \subseteq \mathcal{M}^-}{\text{maximize}} && t_w(G_{E_{\text{init}} \setminus E}) \\ & \text{subject to} && |E| = k. \end{aligned} \quad (15)$$

Remark 1. It is easy to see that every instance of $k\text{-ESP}^-$ can be expressed as an instance of $d\text{-ESP}^+$ problem for a different base graph, some d and a candidate set \mathcal{M}^+ (and vice versa).

Remark 2. The open problem of identifying t -optimal graphs among all graphs with n vertices and m edges [4] is an instance of $k\text{-ESP}^+$ with $k = m$, $E_{\text{init}} = \emptyset$ and $\mathcal{M}^+ = E(K_n)$.

Remarks 1 and 2 ensure that any algorithm designed for solving $k\text{-ESP}^+$ carries over to the other forms of ESP. Therefore, although many graph sparsification and edge pruning scenarios can be naturally stated as a $k\text{-ESP}^-$, in the rest of this paper we focus our attention mainly on $k\text{-ESP}^+$.

4.2 Exhaustive Search

The brute force algorithm for solving $k\text{-ESP}^+$ requires computing the weighted tree-connectivity of every k -subset of the candidate set. $t_w(G)$ can be computed by performing a Cholesky decomposition on the reduced weighted Laplacian matrix which requires $\mathcal{O}(n^3)$ time in general. This time may significantly reduce for sparse graphs. Let $c \triangleq |\mathcal{M}^+|$. For $k = \mathcal{O}(1)$, the time complexity of the brute force algorithm is $\mathcal{O}(c^k n^3)$. If $c = \mathcal{O}(n^2)$, this complexity becomes $\mathcal{O}(n^{2k+3})$, which clearly is not scalable beyond $k \geq 3$. Moreover, for $k = \alpha \cdot c$ ($\alpha < 1$) the time complexity of exhaustive search becomes exponential in c . To address this problem, in the rest of this section we propose two efficient approximation algorithms with performance guarantees by exploiting the inherent structures of tree-connectivity.

4.3 Greedy Algorithm

For any $n \geq 2$, $w : E(K_n) \rightarrow \mathbb{R}_{>0}$, connected (V, E_{init}) , and $\mathcal{M}^+ \subseteq E(K_n)$ define

$$\varphi : 2^{\mathcal{M}^+} \rightarrow \mathbb{R}_{\geq 0} \quad (16)$$

$$E \mapsto \log \text{TG}_{n,w}(E) \quad (17)$$

Note that φ is essentially $\log \text{TG}_{n,w}$ restricted to \mathcal{M}^+ . Therefore, Corollary 4.1 readily follows from Theorem 3.4.

Corollary 4.1. *φ is normalized, monotone and submodular.*

Consequently, $k\text{-ESP}^+$ can be expressed as the problem of maximizing a normalized monotone submodular function subject to a cardinality constraint, i.e.,

$$\begin{aligned} & \underset{E \subseteq \mathcal{M}^+}{\text{maximize}} && \varphi(E) \\ & \text{subject to} && |E| = k. \end{aligned} \quad (18)$$

Maximizing an arbitrary monotone submodular function subject to a cardinality constraint *can be* NP-hard in general (see e.g., the Maximum Coverage problem [13]). Therefore it is reasonable to look for reliable approximation algorithms. In this section we study the greedy algorithm described in Algorithm 1. Theorem 4.2 guarantees that Algorithm 1 is a constant-factor approximation algorithm for $k\text{-ESP}^+$ with a factor of $(1 - 1/e) \approx 0.63$.

Theorem 4.2 (Nemhauser et al. [23]). *The greedy algorithm attains at least $(1 - 1/e)f^*$, where f^* is the maximum of any normalized monotone submodular function subject to a cardinality constraint.*²

Remark 3. Recall that φ is normalized by $\log \text{tree}_{n,w}(E_{\text{init}})$, and therefore reflects the *tree-connectivity gain* achieved by adding k new edges to the original graph (V, E_{init}, w) . In order to avoid any confusion, from now on we denote the optimum value of (18) by OPT^φ , and use OPT to refer to the maximum achievable tree-connectivity in $k\text{-ESP}^+$. Note that,

$$\text{OPT}^\varphi = \text{OPT} - \log \text{tree}_{n,w}(E_{\text{init}}). \quad (19)$$

²A generalized version of Theorem 4.2 [17] states that after $\ell \geq k$ steps, the greedy algorithm is guaranteed to achieve at least $(1 - e^{-\ell/k})f_k^*$, where f_k^* is the maximum of $f(\mathcal{A})$ subject to $|\mathcal{A}| = k$.

Algorithm 1 Greedy Edge Selection

```

1: function GreedyESP( $\mathbf{L}_{\text{init}}, \mathcal{M}^+, k$ )
2:    $E \leftarrow \emptyset$ 
3:    $\mathbf{L} \leftarrow \mathbf{L}_{\text{init}}$ 
4:    $\mathbf{C} \leftarrow \text{Cholesky}(\mathbf{L})$ 
5:   while  $|E| < k$  do
6:      $e_{uv}^* \leftarrow \text{BestEdge}(\mathcal{M}^+ \setminus E, \mathbf{C})$ 
7:      $E \leftarrow E \cup \{e^*\}$ 
8:      $\mathbf{a}_{uv} \leftarrow \mathbf{e}_u - \mathbf{e}_v$ 
9:      $\mathbf{L} \leftarrow \mathbf{L} + w(e_{uv}^*) \mathbf{a}_{uv} \mathbf{a}_{uv}^\top$ 
10:     $\mathbf{C} \leftarrow \text{CholeskyUpdate}(\mathbf{C}, \sqrt{w(e_{uv}^*)} \mathbf{a}_{uv})$  ▷ Rank-one update
11:   end while
12:   return  $E$ 
13: end function

```

```

14: function BestEdge( $\mathcal{M}, \mathbf{C}$ )
15:    $m \leftarrow 0$  ▷ Maximum value
16:   for all  $e \in \mathcal{M}$  do ▷ Parallelizable loop
17:      $w_e \leftarrow w(e)$ 
18:      $\Delta_e \leftarrow \text{Reff}(e, \mathbf{C})$ 
19:     if  $w_e \Delta_e > m$  then
20:        $e^* \leftarrow e$ 
21:        $m \leftarrow w_e \Delta_e$ 
22:     end if
23:   end for
24:   return  $e^*$ 
25: end function

```

```

26: function Reff( $e_{uv}, \mathbf{C}$ ) ▷ Effective Resistance
27:    $\mathbf{a}_{uv} \leftarrow \mathbf{e}_u - \mathbf{e}_v$ 
28:   // solve  $\mathbf{C} \mathbf{x}_{uv} = \mathbf{a}_{uv}$ 
29:    $\mathbf{x}_{uv} \leftarrow \text{ForwardSolver}(\mathbf{C}, \mathbf{a}_{uv})$  ▷ Lower Triangular
30:    $\Delta_{uv} \leftarrow \|\mathbf{x}_{uv}\|^2$ 
31:   return  $\Delta_{uv}$ 
32: end function

```

Let E_{greedy} be the set of edges picked by Algorithm 1. Define $\varphi_{\text{greedy}} \triangleq \varphi(E_{\text{greedy}})$. Then, according to Theorem 4.2, $\varphi_{\text{greedy}} \geq (1 - 1/e) \text{OPT}^\varphi$ and therefore,

$$\log \text{tree}_{n,w}(E_{\text{greedy}} \cup E_{\text{init}}) \geq (1 - 1/e) \text{OPT} + 1/e \log \text{tree}_{n,w}(E_{\text{init}}). \quad (20)$$

Algorithm 1 starts with an empty set of edges, and in each round picks the edge that maximizes the weighted tree-connectivity of the graph, until the cardinality requirement is met. Hence now we need a procedure for finding the edge that maximizes the weighted tree-connectivity. An efficient strategy is to use Lemma 3.1 and pick the edge with the highest effective resistance $w_{uv} \Delta_{uv}$. To compute $\Delta_{uv} = \mathbf{a}_{uv}^\top \mathbf{L}^{-1} \mathbf{a}_{uv}$, we first compute the Cholesky factor of the reduced weighted Laplacian matrix of the current graph $\mathbf{L} = \mathbf{C} \mathbf{C}^\top$. Next, we note that $\Delta_{uv} = \|\mathbf{x}_{uv}\|^2$ where \mathbf{x}_{uv} is the solution of the triangular system $\mathbf{C} \mathbf{x}_{uv} = \mathbf{a}_{uv}$. \mathbf{x}_{uv} can be computed by forward substitution in $\mathcal{O}(n^2)$ time. The time complexity of each round is dominated by the

$\mathcal{O}(n^3)$ time required for computing the Cholesky factor \mathbf{C} . In the i th round, Algorithm 1 has to compute $c-i$ effective resistances where $c = |\mathcal{M}^+|$. For $k = \alpha \cdot c$ ($\alpha < 1$), evaluating effective resistances takes $\mathcal{O}(c^2 n^2)$ time. If $k = \mathcal{O}(1)$, this time reduces to $\mathcal{O}(c n^2)$. Also, note that upon computing the Cholesky factor once in each round, \mathbf{x}_{uv} 's can be computed in parallel by solving $\mathbf{C}\mathbf{x}_{uv} = \mathbf{a}_{uv}$ for different values of \mathbf{a}_{uv} (see line #16 in Algorithm 1). We can avoid the $\mathcal{O}(k n^3)$ time spent on repetitive Cholesky factorization by factorizing \mathbf{L}_{init} once, followed by $k-1$ rank-one updates, each of which takes $\mathcal{O}(n^2)$ time. Therefore, the total time complexity of Algorithm 1 for $k = \mathcal{O}(1)$ and $k = \alpha \cdot c$ will be $\mathcal{O}(n^3 + c n^2)$ and $\mathcal{O}(n^3 + c^2 n^2)$, respectively. In the worst case of $\mathcal{M}^+ = E(K_n)$, $c = \mathcal{O}(n^2)$ and therefore we get $\mathcal{O}(n^4)$ and $\mathcal{O}(n^6)$, respectively, for $k = \mathcal{O}(1)$ and $k = \alpha \cdot c$. Finally, note that for sparse graphs this complexity drops significantly given a sufficiently good fill-reducing permutation for the reduced weighted graph Laplacian.

4.4 Convex Relaxation

Now we take a different approach and design an efficient approximation algorithm for k -ESP⁺ by means of convex relaxation. We begin by assigning an auxiliary variable $0 \leq \pi_i \leq 1$ to each candidate edge $e_i \in \mathcal{M}^+$. Let $\boldsymbol{\pi} \triangleq [\pi_1 \pi_2 \dots \pi_c]^\top$ be the stacked vector of auxiliary variables in which $c = |\mathcal{M}^+|$. Let $G = (V, E_{\text{init}}, w)$ be the given base graph. Define

$$\mathbf{L}(\boldsymbol{\pi}; G, \mathcal{M}^+) \triangleq \sum_{e_i \in E_{\text{init}}} \mathbf{L}_{e_i} + \sum_{e_j \in \mathcal{M}^+} \pi_j \mathbf{L}_{e_j} = \mathbf{A}\mathbf{W}'\mathbf{A}^\top, \quad (21)$$

where \mathbf{L}_{e_k} is the corresponding reduced elementary weighted Laplacian, \mathbf{A} is the reduced incidence matrix of $(V, E_{\text{init}} \cup \mathcal{M}^+)$, and $\mathbf{W}' \triangleq \text{diag}(w'(e_1), \dots, w'(e_s))$ in which $s \triangleq |E_{\text{init}}| + |\mathcal{M}^+|$ and,

$$w'(e_i) \triangleq \begin{cases} \pi_i w(e_i) & e_i \in \mathcal{M}^+, \\ w(e_i) & e_i \notin \mathcal{M}^+. \end{cases} \quad (22)$$

Lemma 4.1. $\mathbf{L}(\boldsymbol{\pi})$ is positive definite iff $(V, E_{\text{init}} \cup \mathcal{M}^+)$ is connected.

Note that every k -subset of \mathcal{M}^+ is optimal for k -ESP⁺ if $(V, E_{\text{init}} \cup \mathcal{M}^+)$ is not connected. Therefore, if we ignore this degenerate case, we can safely assume that $\mathbf{L}(\boldsymbol{\pi}; G, \mathcal{M}^+)$ is positive definite. With a slight abuse of notation, from now on we drop the parameters from $\mathbf{L}(\boldsymbol{\pi}; G, \mathcal{M}^+)$ and use $\mathbf{L}(\boldsymbol{\pi})$ whenever G and \mathcal{M}^+ are clear from the context. Now consider the following optimization problem over $\boldsymbol{\pi}$.

$$\begin{aligned} & \underset{\boldsymbol{\pi}}{\text{maximize}} && \log \det \mathbf{L}(\boldsymbol{\pi}) \\ & \text{subject to} && \|\boldsymbol{\pi}\|_0 = k, \\ & && 0 \leq \pi_i \leq 1, \forall i \in [c]. \end{aligned} \quad (\text{P}_1)$$

P_1 is equivalent to our former definition of k -ESP⁺. The auxiliary variables act as selectors: the i th candidate edge is selected iff $\pi_i = 1$. The objective function rewards strong weighted tree-connectivity. The combinatorial difficulty of ESP here is embodied in the non-convex ℓ_0 -norm constraint. It is easy to see that

at the optimal solution, auxiliary variables take binary values. Therefore \mathbf{P}_1 can also be expressed as

$$\begin{aligned} & \underset{\boldsymbol{\pi}}{\text{maximize}} && \log \det \mathbf{L}(\boldsymbol{\pi}) \\ & \text{subject to} && \|\boldsymbol{\pi}\|_1 = k, \\ & && \pi_i \in \{0, 1\}, \forall i \in [c]. \end{aligned} \tag{P'_1}$$

A natural choice for relaxing \mathbf{P}'_1 is to replace $\pi_i \in \{0, 1\}$ with $0 \leq \pi_i \leq 1$, i.e.,

$$\begin{aligned} & \underset{\boldsymbol{\pi}}{\text{maximize}} && \log \det \mathbf{L}(\boldsymbol{\pi}) \\ & \text{subject to} && \|\boldsymbol{\pi}\|_1 = k, \\ & && 0 \leq \pi_i \leq 1, \forall i \in [c]. \end{aligned} \tag{P_2}$$

The feasible set of \mathbf{P}_2 contains that of \mathbf{P}_1 (or, equivalently, \mathbf{P}'_1), and therefore the optimum value of \mathbf{P}_2 is an upper bound for the optimum of \mathbf{P}_1 (or, equivalently, \mathbf{P}'_1). Note that the ℓ_1 -norm here is identical to $\sum_{i=1}^c \pi_i$. \mathbf{P}_2 is a convex optimization problem since the objective function (tree-connectivity) is concave and the constraints are linear and affine in $\boldsymbol{\pi}$. In fact, \mathbf{P}_2 is an instance of the MAXDET problem [29] subject to additional affine constraints on $\boldsymbol{\pi}$. It is worth noting that \mathbf{P}_2 can be reached also by relaxing the non-convex ℓ_0 -norm constraint in \mathbf{P}_1 by a convex ℓ_1 -norm constraint (essentially $\sum_{i=1}^c \pi_i = k$). Furthermore, \mathbf{P}_2 is also closely related to a ℓ_1 -regularalized instance of MAXDET,

$$\begin{aligned} & \underset{\boldsymbol{\pi}}{\text{maximize}} && \log \det \mathbf{L}(\boldsymbol{\pi}) - \lambda \|\boldsymbol{\pi}\|_1 \\ & \text{subject to} && 0 \leq \pi_i \leq 1, \forall i \in [c]. \end{aligned} \tag{P_3}$$

This problem is a penalized form of \mathbf{P}_2 ; these two problems are equivalent for some positive value of λ . Problem \mathbf{P}_3 is also a convex optimization problem for non-negative λ . The ℓ_1 -norm in \mathbf{P}_3 encourages sparser $\boldsymbol{\pi}$, while the log-determinant rewards stronger tree-connectivity. The penalty coefficient λ is a parameter that specifies the desired degree of sparsity, i.e., larger λ yields a sparser vector of selectors $\boldsymbol{\pi}$.

Problem \mathbf{P}_2 (and \mathbf{P}_3) can be solved efficiently using interior-point methods [5]. After finding a globally optimal solution $\boldsymbol{\pi}^*$ for the relaxed problem \mathbf{P}_2 , we ultimately need to map it into a feasible $\boldsymbol{\pi}$ for \mathbf{P}_1 , i.e., picking k edges from the candidate set \mathcal{M}^+ . First note that if $\boldsymbol{\pi}^* \in \{0, 1\}^c$, it means that $\boldsymbol{\pi}^*$ is already an optimal solution for k -ESP⁺ and \mathbf{P}_1 . However, in the more likely case of $\boldsymbol{\pi}^*$ containing fractional values, we need a *rounding procedure* to set k auxiliary variables to one and others to zero. The most intuitive choice is to pick the k edges with the largest π_i^* 's. Another (approximate) rounding strategy (and a justification for picking the k largest π_i^*) emerges from interpreting π_i as the probability of selecting the i th candidate edge. Theorem 4.3 provides a new interesting way of interpreting the convex relaxation of \mathbf{P}_1 by \mathbf{P}_2 .

Theorem 4.3. Define $E_\bullet \triangleq E_{\text{init}} \cup \mathcal{M}^+$ and $G_\bullet \triangleq (V, E_\bullet, w)$. Let $\boldsymbol{\pi}_\bullet = [\pi_1 \dots \pi_s]^\top \in (0, 1]^s$ such that $s \triangleq |E_{\text{init}}| + |\mathcal{M}^+|$ and $\pi_i = 1$ if $e_i \in E_{\text{init}}$. Then we have

$$\mathbb{E}_{\mathcal{H} \sim \mathbb{G}(G_\bullet, \boldsymbol{\pi}_\bullet)} [t_w(\mathcal{H})] = \det \mathbf{L}(\boldsymbol{\pi}), \tag{23}$$

$$\mathbb{E}_{\mathcal{H} \sim \mathbb{G}(G_\bullet, \boldsymbol{\pi}_\bullet)} [|E(\mathcal{H})| - |E_{\text{init}}|] = \sum_{e_i \in \mathcal{M}^+} \pi_i = \|\boldsymbol{\pi}\|_1. \tag{24}$$

Note that (23) and (24) appear in the objective function and the constraints of \mathbf{P}_2 , respectively. Thus \mathbf{P}_2 can be rewritten as

$$\begin{aligned} & \underset{\pi}{\text{maximize}} && \mathbb{E}_{\mathcal{H} \sim \mathbb{G}(G_\bullet, \pi_\bullet)}[t_w(\mathcal{H})] \\ & \text{subject to} && \mathbb{E}_{\mathcal{H} \sim \mathbb{G}(G_\bullet, \pi_\bullet)}[|E(\mathcal{H})|] = k + |E_{\text{init}}|, \\ & && 0 \leq \pi_i \leq 1, \forall i \in [s]. \end{aligned} \tag{P'_2}$$

This offers a new narrative: the objective in \mathbf{P}_2 is to find the optimal probabilities π^* for sampling edges from \mathcal{M}^+ such that the weighted number of spanning trees is maximized in *expectation*, while the *expected* number of newly selected edges is equal to k . In other words, \mathbf{P}_2 can be seen as a convex relaxation of \mathbf{P}_1 at the expense of maximizing the objective and satisfying the constraint, both *in expectation*. This new interpretation motivates an approximate randomized rounding procedure that picks $e_i \in \mathcal{M}^+$ with probability π_i^* . According to Theorem 4.3, this randomized rounding scheme, in average, attains $\det \mathbf{L}(\pi^*)$ by picking k new edges in average.

Theorem 4.4. *For any $0 < \epsilon < 1$ and $\delta > 0$,*

$$\mathbb{P}[|E^*| < (1 - \epsilon)k] < \exp(-\epsilon^2 k/2), \tag{25}$$

$$\mathbb{P}[|E^*| > (1 + \delta)k] < \exp(-\delta^2 k/3), \tag{26}$$

where E^* is the set of selected edges by the randomized rounding scheme defined above.

Theorem 4.4 ensures that the probability of the events in which the aforementioned randomized rounding strategy picks too many/few edges (compared to k) decay exponentially. Note that this new narrative offers another intuitive justification for deterministically picking the k edges with largest π_i^* 's. Finally, we believe that Theorems 4.3 and 4.4 can potentially be used as building blocks to design new randomized rounding schemes.

4.5 Certifying Near-Optimality

The proposed approximation algorithms also provide *a posteriori* lower and upper bounds for the maximum achievable tree-connectivity in ESP. Let E_{greedy} , E_{cvx} be the solutions returned by the greedy and convex³ approximation algorithms, respectively. Let τ_{cvx}^* be the optimum value of \mathbf{P}_2 and define $\tau_{\text{init}} \triangleq \log \text{tree}_{n,w}(E_{\text{init}})$, $\tau_{\text{cvx}} \triangleq \log \text{tree}_{n,w}(E_{\text{cvx}} \cup E_{\text{init}})$ and $\tau_{\text{greedy}} \triangleq \log \text{tree}_{n,w}(E_{\text{greedy}} \cup E_{\text{init}})$.

Corollary 4.5.

$$\max \left\{ \tau_{\text{greedy}}, \tau_{\text{cvx}} \right\} \leq \text{OPT} \leq \min \left\{ \zeta \tau_{\text{greedy}} + (1 - \zeta) \tau_{\text{init}}, \tau_{\text{cvx}}^* \right\} \tag{27}$$

where $\zeta \triangleq (1 - 1/e)^{-1} \approx 1.58$.⁴

Corollary 4.5 can be used as a tool to assess the quality of any suboptimal design. Let \mathcal{A} be an arbitrary k -subset of \mathcal{M}^+ and $\tau_{\mathcal{A}} = \log \text{tree}_{n,w}(\mathcal{A} \cup E_{\text{init}})$. Define $\mathcal{U} \triangleq \min \left\{ \zeta \tau_{\text{greedy}} + (1 - \zeta) \tau_{\text{init}}, \tau_{\text{cvx}}^* \right\}$. \mathcal{U} can be computed by running the proposed greedy and convex approximation algorithms. From Corollary 4.5 it readily follows that $\text{OPT} - \tau_{\mathcal{A}} \leq \mathcal{U} - \tau_{\mathcal{A}}$ and $\text{OPT}/\tau_{\mathcal{A}} \leq \mathcal{U}/\tau_{\mathcal{A}}$. Therefore, although we may not have

³Picking the k edges with the largest π_i^* 's from the solution of \mathbf{P}_2 .

⁴Furthermore, recall that the leftmost term in (27) is bounded from below by the expression given in (20).

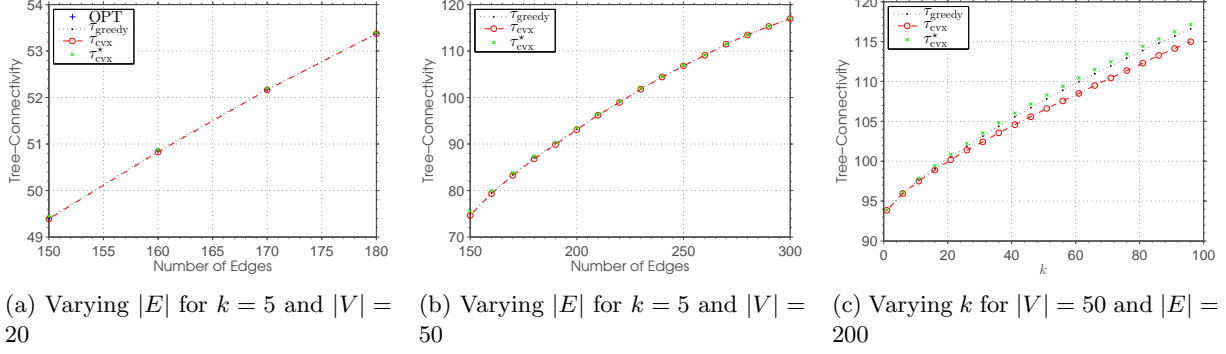


Figure 1: ESP on randomly generated graphs. Recall that according to Corollary 4.5, $\tau_{\text{greedy}} \leq \text{OPT} \leq \tau_{\text{cvx}}^*$.

direct access to OPT, we can still certify the near-optimality of any design such as \mathcal{A} whose $\delta \triangleq \mathcal{U} - \tau_{\mathcal{A}}$ is sufficiently small.

4.6 Numerical Results

We implemented Algorithm 1 in MATLAB. Problem P_2 is modelled using CVX [12, 11] and YALMIP [18], and solved using SDPT3 [28]. Figure 1 illustrates the performance of our approximate solutions to $k\text{-ESP}^+$ in randomly generated graphs. The search space in these experiments is $\mathcal{M}^+ = E(K_n) \setminus E_{\text{init}}$. Figures 1a and 1b show tree-connectivity as a function of number of randomly generated edges for a fixed $k = 5$ and, respectively, $|V| = 20$ and $|V| = 50$. Our results indicate that both algorithms exhibit remarkable performances for $k = 5$. Note that computing OPT by exhaustive search is only feasible in small instances such as Figure 1a. Nevertheless, computing the exact OPT is not crucial for evaluating our approximate algorithms, as it is tightly bounded in $[\tau_{\text{greedy}}, \tau_{\text{cvx}}^*]$ as predicted by Corollary 4.5 (i.e., between each black \cdot and green \times). Figure 1c shows the results obtained for varying k . The optimality gap for τ_{cvx} gradually grows as the planning horizon k increases. Our greedy algorithm, however, still yields a near-optimal approximation.

5 Beyond $k\text{-ESP}^+$

5.1 Matroid Constraints

Recall that φ is monotone. Therefore, except the degenerate case of $(V, E_{\text{init}} \cup \mathcal{M}^+)$ not being connected, replacing the cardinality constraint $|E| = k$ in $k\text{-ESP}^+$ with an inequality constraint $|E| \leq k$ does not affect the set of optimal solutions. Consider the *uniform matroid* [24] defined as $(\mathcal{M}^+, \mathcal{I}_U)$ where

$$\mathcal{I}_U \triangleq \{\mathcal{A} \subseteq \mathcal{M}^+ : |\mathcal{A}| \leq k\}.$$

The inequality cardinality constraint can be expressed as $E \in \mathcal{I}_U$.

Definition 5.1 (Partition Matroid). Let $\mathcal{M}_1^+, \dots, \mathcal{M}_\ell^+$ be a partition for \mathcal{M}^+ . Assign an integer (budget)

$0 \leq k_i \leq |\mathcal{M}_i^+|$ to each \mathcal{M}_i^+ . Define

$$\mathcal{I}_P \triangleq \left\{ \mathcal{A} \subseteq \mathcal{M}^+ : |\mathcal{A} \cap \mathcal{M}_i^+| \leq k_i \text{ for } i \in [\ell] \right\}.$$

The pair $(\mathcal{M}^+, \mathcal{I}_P)$ is called a *partition matroid*.

Now let us consider ESP under a partition matroid constraint; i.e.,

$$\begin{aligned} & \text{maximize} && \varphi(E) \\ & \text{subject to} && E \in \mathcal{I}_P. \end{aligned} \tag{28}$$

Note that k -ESP⁺ is a special case of this problem with $\ell = 1$ and $k_1 = k$. Now, by choosing different partitions for \mathcal{M}^+ and different budgets k_i we can model a wide variety of graph synthesis problems. For example consider the following extension of k -ESP⁺,

$$\begin{aligned} & \text{maximize}_{E \subseteq \mathcal{M}^+, |E| \leq k} && \varphi(E) \\ & \text{subject to} && \deg(v) \leq d. \end{aligned} \tag{29}$$

Define $\mathcal{M}_v^+ \triangleq \{e \in \mathcal{M}^+ : v \in e\}$. Now note that the constraints in (29) can be expressed as a partition matroid with two blocks: (i) \mathcal{M}_v^+ with a budget of $k_1 = d$, and (ii) $\mathcal{M}^+ \setminus \mathcal{M}_v^+$ with a budget of $k_2 = k - d$.

5.1.1 Greedy Algorithm

Theorem 5.1 (Fisher et al. [7]). *The greedy algorithm attains at least $(1/2)f^*$, where f^* is the maximum of any normalized monotone submodular function subject to a matroid constraint.*

According to Theorem 5.1, a slightly modified version of Algorithm 1, that abides by the matroid constraint while greedily choosing the next best edge, yields a $\frac{1}{2}$ -approximation [7, 17].

5.1.2 Convex Relaxation

The proposed convex relaxation of k -ESP⁺ can be modified to handle a partition matroid constraint. First note that (28) can be expressed as

$$\begin{aligned} & \text{maximize}_{\boldsymbol{\pi}} && \log \det \mathbf{L}(\boldsymbol{\pi}) \\ & \text{subject to} && \sum_{e_i \in \mathcal{M}_j^+} \pi_i \leq k_j, \forall j \in [\ell] \\ & && \pi_i \in \{0, 1\}, \forall i \in [c]. \end{aligned} \tag{P_4}$$

Relaxing the binary constraints on π_i 's yields

$$\begin{aligned}
& \underset{\boldsymbol{\pi}}{\text{maximize}} && \log \det \mathbf{L}(\boldsymbol{\pi}) \\
& \text{subject to} && \sum_{e_i \in \mathcal{M}_j^+} \pi_i \leq k_j, \forall j \in [\ell] \\
& && 0 \leq \pi_i \leq 1, \forall i \in [c].
\end{aligned} \tag{P_5}$$

\mathbf{P}_5 is a convex optimization problem and, as before, can be solved efficiently using interior-point methods. A simple rounding strategy for the solution of \mathbf{P}_5 is to pick the edges in \mathcal{M}_i^+ that are associated to the k_i largest π_j^* 's (for $i \in [\ell]$). Moreover, the bounds in (27) (with $\zeta = 2$) and Theorem 4.3 can also be readily generalized to handle partition matroid constraints. In particular, the optimum value of \mathbf{P}_5 gives an upper bound for the optimum value of \mathbf{P}_4 . Also, similar to Theorem 4.3, \mathbf{P}_5 can be interpreted as maximizing the expected value of the weighted number of spanning trees such that the expected number of new edges sampled from \mathcal{M}_i^+ is at most k_i , for $i \in [\ell]$.

5.2 Dual of k -ESP⁺

The dual of k -ESP⁺ aims to identify and select the minimal set of new edges from a candidate set \mathcal{M}^+ such that the resulting tree-connectivity gain is at least $0 \leq \delta \leq \varphi(\mathcal{M}^+)$ for some given δ ; i.e.,

$$\begin{aligned}
& \underset{E \subseteq \mathcal{M}^+}{\text{minimize}} && |E| \\
& \text{subject to} && \varphi(E) \geq \delta.
\end{aligned} \tag{30}$$

5.2.1 Greedy Algorithm

The greedy algorithm for approximating the solution of (30) is outlined in Algorithm 2. The only difference between Algorithm 1 and Algorithm 2 is that the latter terminates when the δ -bound is achieved (or, alternatively, when there are no more edges left in \mathcal{M}^+ , which indicates an empty feasible set). Wolsey [31] proves several upper bounds for the ratio between the objective value achieved by the greedy algorithm and the optimum value of the following class of problems,

$$\underset{\mathcal{A} \subseteq \mathcal{W}}{\text{minimize}} |\mathcal{A}| \text{ subject to } \phi(\mathcal{A}) \geq \phi_0, \tag{31}$$

in which $\phi : 2^{\mathcal{W}} \rightarrow \mathbb{R}$ is an arbitrary monotone submodular function and $\phi_0 \leq \phi(\mathcal{W})$. Note that our problem (30) is special case of (31), and therefore (some of) the bounds proved by Wolsey [31, Theorem 1] also hold for Algorithm 2.

Theorem 5.2 (Wolsey [31]). *Let k_{OPT} and k_{greedy} be the global minimum of (30) and the objective value achieved by Algorithm 2, respectively. Also, let $\tilde{E}_{\text{greedy}}$ be the set formed by Algorithm 2 one step before termination. Then $k_{\text{greedy}} \leq \gamma k_{\text{OPT}}$ in which*

$$\gamma \triangleq 1 + \log \left(\frac{\delta}{\delta - \varphi(\tilde{E}_{\text{greedy}})} \right). \tag{32}$$

Algorithm 2 Greedy Dual Edge Selection

```

1: function GreedyDualESP( $\mathbf{L}_{\text{init}}, \mathcal{M}^+, \delta$ )
2:    $E \leftarrow \emptyset$ 
3:    $\mathbf{L} \leftarrow \mathbf{L}_{\text{init}}$ 
4:    $\mathbf{C} \leftarrow \text{Cholesky}(\mathbf{L})$ 
5:   while  $(\log \det \mathbf{L} < \delta) \wedge (E \neq \mathcal{M}^+)$  do
6:      $e_{uv}^* \leftarrow \text{BestEdge}(\mathcal{M}^+ \setminus E, \mathbf{C})$ 
7:      $E \leftarrow E \cup \{e^*\}$ 
8:      $\mathbf{a}_{uv} \leftarrow \mathbf{e}_u - \mathbf{e}_v$ 
9:      $\mathbf{L} \leftarrow \mathbf{L} + w(e_{uv}^*) \mathbf{a}_{uv} \mathbf{a}_{uv}^\top$ 
10:     $\mathbf{C} \leftarrow \text{CholeskyUpdate}(\mathbf{C}, \sqrt{w(e_{uv}^*)} \mathbf{a}_{uv})$  ▷ Rank-one update
11:   end while
12:   return  $E$ 
13: end function

```

The upper bound given above and some of the other bounds in [31] are *a posteriori* in the sense that they can be computed only *after* running the greedy algorithm.

5.2.2 Convex Relaxation

Let $\tau_{\text{init}} \triangleq \log \det \mathbf{L}(\mathbf{0})$. The dual problem can be expressed as

$$\begin{aligned}
& \underset{\boldsymbol{\pi}}{\text{minimize}} && \sum_{i=1}^c \pi_i \\
& \text{subject to} && \log \det \mathbf{L}(\boldsymbol{\pi}) \geq \delta + \tau_{\text{init}}, \\
& && \pi_i \in \{0, 1\}, \forall i \in [c].
\end{aligned} \tag{D1}$$

The combinatorial difficulty of the dual formulation of ESP is manifested in the binary constraints of \mathbf{D}_1 . Relaxing these constraints into $0 \leq \pi_i \leq 1$ yields the following convex optimization problem,

$$\begin{aligned}
& \underset{\boldsymbol{\pi}}{\text{minimize}} && \sum_{i=1}^c \pi_i \\
& \text{subject to} && \log \det \mathbf{L}(\boldsymbol{\pi}) \geq \delta + \tau_{\text{init}}, \\
& && 0 \leq \pi_i \leq 1, \forall i \in [c].
\end{aligned} \tag{D2}$$

\mathbf{D}_2 can be solved efficiently using interior-point methods. Let $\boldsymbol{\pi}^*$ be the minimizer of \mathbf{D}_2 . $\sum_{i=1}^c \pi_i^*$ is a lower bound for the optimum value of the dual ESP \mathbf{D}_1 . If $\boldsymbol{\pi}^* \in \{0, 1\}^c$, $\boldsymbol{\pi}^*$ is also a globally optimal solution for \mathbf{D}_1 . Otherwise we need a rounding scheme to map $\boldsymbol{\pi}^*$ into a feasible (suboptimal) solution for \mathbf{D}_1 . A simple deterministic rounding strategy is the following.

- Step 1. Sort the edges in \mathcal{M}^+ according to $\boldsymbol{\pi}^*$ in descending order.
- Step 2. Pick edges from the sorted list until $\log \det \mathbf{L}(\boldsymbol{\pi}) \geq \delta + \tau_{\text{init}}$.

Theorem 4.3 allows us to interpret D_2 as finding the optimal sampling probabilities π^* that minimizes the expected number of new edges such that the expected weighted number of spanning trees is at least $\exp(\delta + \tau_{\text{init}})$; i.e.,

$$\begin{aligned} & \underset{\pi}{\text{minimize}} && \mathbb{E}_{\mathcal{H} \sim \mathbb{G}(G_\bullet, \pi_\bullet)}[|E(\mathcal{H})|], \\ & \text{subject to} && \mathbb{E}_{\mathcal{H} \sim \mathbb{G}(G_\bullet, \pi_\bullet)}[t_w(\mathcal{H})] \geq \exp(\delta + \tau_{\text{init}}), \\ & && 0 \leq \pi_i \leq 1, \forall i \in [s], \end{aligned} \tag{D'_2}$$

in which $\mathbb{G}(G_\bullet, \pi_\bullet)$ is defined in Theorem 4.3. This narrative suggests a randomized rounding scheme in which $e_i \in \mathcal{M}^+$ is selected with probability π_i^* . The expected number of selected edges by this procedure is $\sum_{i=1}^c \pi_i^*$.

5.2.3 Certifying Near-Optimality

Corollary 5.3. Define $\zeta^* \triangleq 1/\gamma$ where γ is the approximation factor given by Theorem 5.2. Let k_{cvx} be the number of new edges selected by the deterministic rounding procedure described above.

$$\max \left\{ \zeta^* k_{\text{greedy}}, \left\lceil \sum_{i=1}^c \pi_i^* \right\rceil \right\} \leq k_{\text{OPT}} \leq \min \left\{ k_{\text{greedy}}, k_{\text{cvx}} \right\}. \tag{33}$$

As we did before for $k\text{-ESP}^+$, the lower bound provided by Corollary 5.3 can be used to construct an upper bound for the gap between k_{OPT} and any (feasible) suboptimal design with an objective value of $k_{\mathcal{A}}$. Let $\mathcal{L} \triangleq \max \left\{ \zeta^* k_{\text{greedy}}, \left\lceil \sum_{i=1}^c \pi_i^* \right\rceil \right\}$. \mathcal{L} can be computed by running Algorithm 2 and solving the convex optimization problem D_2 . Consequently, $k_{\mathcal{A}} - k_{\text{OPT}} \leq k_{\mathcal{A}} - \mathcal{L}$ and $k_{\mathcal{A}}/k_{\text{OPT}} \leq k_{\mathcal{A}}/\mathcal{L}$.

6 Conclusion

We studied the problem of designing near- t -optimal graphs under several types of constraints and formulations. Several new structures were revealed and exploited to design efficient approximation algorithms. In particular, we proved that the weighted number of spanning trees in connected graphs can be posed as a monotone log-submodular function of the edge set. Our approximation algorithms can find near-optimal solutions with performance guarantees. They also provide *a posteriori* near-optimality certificates for arbitrary designs. Our results can be readily applied to a wide verity of applications involving graph synthesis and graph sparsification scenarios.

References

- [1] Rosemary A Bailey and Peter J Cameron. Combinatorics of optimal designs. *Surveys in Combinatorics*, 365:19–73, 2009. 3
- [2] Prabir Barooah and Joao P Hespanha. Estimation on graphs from relative measurements. *Control Systems, IEEE*, 27(4):57–74, 2007. 6

- [3] Douglas Bauer, Francis T Boesch, Charles Suffel, and R Van Slyke. On the validity of a reduction of reliable network design to a graph extremal problem. *Circuits and Systems, IEEE Transactions on*, 34(12):1579–1581, 1987. 3
- [4] Francis T Boesch, Appajosyula Satyanarayana, and Charles L Suffel. A survey of some network reliability analysis and synthesis results. *Networks*, 54(2):99–107, 2009. 3, 7
- [5] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004. 11
- [6] Ching-Shui Cheng. Maximizing the total number of spanning trees in a graph: two related problems in graph theory and optimum design theory. *Journal of Combinatorial Theory, Series B*, 31(2):240–248, 1981. 3
- [7] Marshall L Fisher, George L Nemhauser, and Laurence A Wolsey. *An analysis of approximations for maximizing submodular set functionsII*. Springer, 1978. 14
- [8] N Gaffke. D-optimal block designs with at most six varieties. *Journal of Statistical Planning and Inference*, 6(2):183–200, 1982. 3
- [9] Arpita Ghosh, Stephen Boyd, and Amin Saberi. Minimizing effective resistance of a graph. *SIAM review*, 50(1):37–66, 2008. 6
- [10] Chris Godsil and Gordon Royle. *Algebraic graph theory*. Graduate Texts in Mathematics Series. Springer London, Limited, 2001. ISBN 9780387952413. 3, 4
- [11] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. http://stanford.edu/~boyd/graph_dcp.html. 13
- [12] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. <http://cvxr.com/cvx>, March 2014. 13
- [13] Dorit S Hochbaum. *Approximation algorithms for NP-hard problems*. PWS Publishing Co., 1996. 8
- [14] Siddharth Joshi and Stephen Boyd. Sensor selection via convex optimization. *Signal Processing, IEEE Transactions on*, 57(2):451–462, 2009. 23
- [15] Alexander K Kelmans. On graphs with the maximum number of spanning trees. *Random Structures & Algorithms*, 9(1-2):177–192, 1996. 3
- [16] Alexander K Kelmans and BN Kimelfeld. Multiplicative submodularity of a matrix’s principal minor as a function of the set of its rows and some combinatorial applications. *Discrete Mathematics*, 44(1):113–116, 1983. 3
- [17] Andreas Krause and Daniel Golovin. Submodular function maximization. *Tractability: Practical Approaches to Hard Problems*, 3:19, 2012. 8, 14

- [18] Johan Löfberg. Yalmip : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004. URL <http://users.isy.liu.se/johanl/yalmip>. 13
- [19] László Lovász. Random walks on graphs: A survey. *Combinatorics, Paul Erdős is eighty*, 2(1):1–46, 1993. 6
- [20] Mehran Mesbahi and Magnus Egerstedt. *Graph theoretic methods in multiagent networks*. Princeton University Press, 2010. 5
- [21] Carl D Meyer. *Matrix analysis and applied linear algebra*. SIAM, 2000. 20
- [22] Wendy Myrvold. Reliable network synthesis: Some recent developments. In *Proceedings of International Conference on Graph Theory, Combinatorics, Algorithms, and Applications*, 1996. 3
- [23] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions - I. *Mathematical Programming*, 14(1):265–294, 1978. 8
- [24] James G Oxley. *Matroid theory*, volume 3. Oxford university press, 2006. 13
- [25] Louis Petingi and Jose Rodriguez. A new technique for the characterization of graphs with a maximum number of spanning trees. *Discrete mathematics*, 244(1):351–373, 2002. 3
- [26] Friedrich Pukelsheim. *Optimal design of experiments*, volume 50. SIAM, 1993. 3
- [27] DR Shier. Maximizing the number of spanning trees in a graph with n nodes and m edges. *Journal Research National Bureau of Standards, Section B*, 78:193–196, 1974. 3
- [28] Reha H Tütüncü, Kim C Toh, and Michael J Todd. Solving semidefinite-quadratic-linear programs using sdpt3. *Mathematical programming*, 95(2):189–217, 2003. 13
- [29] Lieven Vandenbergh, Stephen Boyd, and Shao-Po Wu. Determinant maximization with linear matrix inequality constraints. *SIAM journal on matrix analysis and applications*, 19(2):499–533, 1998. 11
- [30] Guy Weichenberg, Vincent WS Chan, and Muriel Médard. High-reliability topological architectures for networks under stress. *Selected Areas in Communications, IEEE Journal on*, 22(9):1830–1845, 2004. 3
- [31] Laurence A Wolsey. An analysis of the greedy algorithm for the submodular set covering problem. *Combinatorica*, 2(4):385–393, 1982. 15, 16

A Proofs

Lemma A.1. *For any $\mathbf{M} \in \mathbb{S}_{>0}^n$ and $\mathbf{N} \in \mathbb{S}_{>0}^n$, $\mathbf{M} \succeq \mathbf{N}$ iff $\mathbf{N}^{-1} \succeq \mathbf{M}^{-1}$.*

Proof. Due to symmetry it suffices to prove that $\mathbf{M} \succeq \mathbf{N} \Rightarrow \mathbf{N}^{-1} \succeq \mathbf{M}^{-1}$. Multiplying both sides of $\mathbf{M} \succeq \mathbf{N}$ by $\mathbf{N}^{-\frac{1}{2}}$ from left and right results in $\mathbf{N}^{-\frac{1}{2}}\mathbf{M}\mathbf{N}^{-\frac{1}{2}} - \mathbf{I} \succeq \mathbf{0}$. Therefore the eigenvalues of $\mathbf{N}^{-\frac{1}{2}}\mathbf{M}\mathbf{N}^{-\frac{1}{2}}$, which are the same as the eigenvalues of $\mathbf{M}^{\frac{1}{2}}\mathbf{N}^{-1}\mathbf{M}^{\frac{1}{2}}$,⁵ are at least 1. Therefore $\mathbf{M}^{\frac{1}{2}}\mathbf{N}^{-1}\mathbf{M}^{\frac{1}{2}} - \mathbf{I} \succeq \mathbf{0}$. Multiplying both sides by $\mathbf{M}^{-\frac{1}{2}}$ from left and right proves the lemma. \square

Lemma A.2 (Matrix Determinant Lemma). *For any non-singular $\mathbf{M} \in \mathbb{R}^{n \times n}$ and $\mathbf{c}, \mathbf{d} \in \mathbb{R}^n$,*

$$\det(\mathbf{M} + \mathbf{c}\mathbf{d}^\top) = (1 + \mathbf{d}^\top \mathbf{M}^{-1} \mathbf{c}) \det \mathbf{M}. \quad (34)$$

Proof. See e.g., [21]. \square

Lemma A.3. *Let G_1 be a spanning subgraph of G_2 . For any $w : E(K) \rightarrow \mathbb{R}_{\geq 0}$, $\mathbf{L}_{G_2}^w \succeq \mathbf{L}_{G_1}^w$ in which \mathbf{L}_G^w is the reduced weighted Laplacian matrix of G when its edges are weighted by w .*

Proof. From the definition of the reduced weighted Laplacian matrix we have,

$$\mathbf{L}_{G_2}^w - \mathbf{L}_{G_1}^w = \sum_{\{u,v\} \in E(G_2) \setminus E(G_1)} w_{uv} \mathbf{a}_{uv} \mathbf{a}_{uv}^\top \succeq \mathbf{0}. \quad (35)$$

\square

Proof of Theorem 3.1. Define the following indicator function,

$$\mathbb{1}_{\mathcal{T}_G}(T) \triangleq \begin{cases} 1 & T \in \mathcal{T}_G, \\ 0 & T \notin \mathcal{T}_G, \end{cases} \quad (36)$$

in which \mathcal{T}_G denotes the set of spanning trees of G . Now note that,

$$\mathbb{E}_{G \sim \mathbb{G}(G^\circ, \mathbf{p})}[t_w(G)] = \mathbb{E}_{G \sim \mathbb{G}(G^\circ, \mathbf{p})} \left[\sum_{T \in \mathcal{T}_{G^\circ}} \mathbb{1}_{\mathcal{T}_G}(T) \mathbb{V}_w(T) \right] \quad (37)$$

$$= \sum_{T \in \mathcal{T}_{G^\circ}} \mathbb{E}_{G \sim \mathbb{G}(G^\circ, \mathbf{p})} \left[\mathbb{1}_{\mathcal{T}_G}(T) \mathbb{V}_w(T) \right] \quad (38)$$

$$= \sum_{T \in \mathcal{T}_{G^\circ}} \mathbb{P}[T \in \mathcal{T}_G] \mathbb{V}_w(T) \quad (39)$$

$$= \sum_{T \in \mathcal{T}_{G^\circ}} \mathbb{V}_p(T) \mathbb{V}_w(T) \quad (40)$$

$$= \sum_{T \in \mathcal{T}_{G^\circ}} \mathbb{V}_{w_p}(T) \quad (41)$$

$$= t_{w_p}(G^\circ). \quad (42)$$

⁵Recall that $\mathbf{M}\mathbf{N}$ and $\mathbf{N}\mathbf{M}$ have the same spectrum.

Here we have used the fact the $\mathbb{P}[T \in \mathcal{T}_G]$ is equal to the probability of existence of every edge of T in G , which is equal to $\mathbb{V}_p(T)$. \square

Proof of Lemma 3.1. Note that $\mathbf{L}_{G^+} = \mathbf{L}_G + w_{uv} \mathbf{a}_{uv} \mathbf{a}_{uv}^\top$. Taking the determinant, applying Lemma A.2 and taking the log concludes the proof. \square

Proof of Lemma 3.2. The proof is similar to the proof of Lemma 3.1. \square

Proof of Theorem 3.3. First recall that $\mathbb{V}_w(T)$ is positive for any T by definition.

1. Normalized: $\text{tree}_{n,w}(\emptyset) = 0$ by definition.
2. Monotone: Let $G \triangleq (V, E \cup \{e\})$. Denote by \mathcal{T}_G^e the set of spanning trees of G that contain e .

$$\text{tree}_{n,w}(E \cup \{e\}) = \sum_{T \in \mathcal{T}_G} \mathbb{V}_w(T) = \sum_{T \in \mathcal{T}_G^e} \mathbb{V}_w(T) + \sum_{T \notin \mathcal{T}_G^e} \mathbb{V}_w(T) \quad (43)$$

$$= \sum_{T \in \mathcal{T}_G^e} \mathbb{V}_w(T) + \text{tree}_{n,w}(E) \geq \text{tree}_{n,w}(E). \quad (44)$$

3. Supermodular: $\text{tree}_{n,w}$ is supermodular iff for all $E_1 \subseteq E_2 \subseteq E(K_n)$ and all $e \in E(K_n) \setminus E_2$,

$$\text{tree}_{n,w}(E_2 \cup \{e\}) - \text{tree}_{n,w}(E_2) \geq \text{tree}_{n,w}(E_1 \cup \{e\}) - \text{tree}_{n,w}(E_1). \quad (45)$$

Define $G_1 \triangleq (V, E_1)$ and $G_2 \triangleq (V, E_2)$. As we showed in (44),

$$\text{tree}_{n,w}(E_1 \cup \{e\}) - \text{tree}_{n,w}(E_1) = \sum_{T \in \mathcal{T}_{G_1}^e} \mathbb{V}_w(T), \quad (46)$$

$$\text{tree}_{n,w}(E_2 \cup \{e\}) - \text{tree}_{n,w}(E_2) = \sum_{T \in \mathcal{T}_{G_2}^e} \mathbb{V}_w(T). \quad (47)$$

Therefore we need to show that $\sum_{T \in \mathcal{T}_{G_2}^e} \mathbb{V}_w(T) \geq \sum_{T \in \mathcal{T}_{G_1}^e} \mathbb{V}_w(T)$. This inequality holds since $\mathcal{T}_{G_1}^e \subseteq \mathcal{T}_{G_2}^e$. \square

Proof of Theorem 3.4.

1. Normalized: By definition $\log \text{TG}_{n,w}(\emptyset) = \log \text{tree}_{n,w}(E_{\text{init}}) - \log \text{tree}_{n,w}(E_{\text{init}}) = 0$.
2. Monotone: We need to show that $\log \text{TG}_{n,w}(E \cup \{e\}) \geq \log \text{TG}_{n,w}(E)$. This is equivalent to showing that,

$$\log \text{tree}_{n,w}(E_{\text{init}} \cup E \cup \{e\}) \geq \log \text{tree}_{n,w}(E_{\text{init}} \cup E). \quad (48)$$

Now note that $(V, E_{\text{init}} \cup E)$ is connected since (V, E_{init}) was assumed to be connected. Therefore we can apply Lemma 3.1 on the LHS of (48); i.e.,

$$\log \text{tree}_{n,w}(E_{\text{init}} \cup E \cup \{e\}) = \log \text{tree}_{n,w}(E_{\text{init}} \cup E) + \log(1 + w_e \Delta_e). \quad (49)$$

Therefore it suffices to show that $\log(1 + w_e \Delta_e)$ is non-negative. Since (V, E_{init}) is connected, \mathbf{L} is positive definite. Consequently $w_e \Delta_e = w_e \mathbf{a}_e^\top \mathbf{L}^{-1} \mathbf{a}_e > 0$ and hence $\log(1 + w_e \Delta_e) > 0$.

3. Submodular: $\log \text{TG}_{n,w}$ is submodular iff for all $E_1 \subseteq E_2 \subseteq E(K_n)$ and all $e \in E(K_n) \setminus E_2$,

$$\log \text{TG}_{n,w}(E_1 \cup \{e\}) - \log \text{TG}_{n,w}(E_1) \geq \log \text{TG}_{n,w}(E_2 \cup \{e\}) - \log \text{TG}_{n,w}(E_2). \quad (50)$$

After canceling $\log \text{tree}_{n,w}(E_{\text{init}})$ we need to show that,

$$\log \text{tree}_{n,w}(E_1 \cup E_{\text{init}} \cup \{e\}) - \log \text{tree}_{n,w}(E_1 \cup E_{\text{init}}) \geq \log \text{tree}_{n,w}(E_2 \cup E_{\text{init}} \cup \{e\}) - \log \text{tree}_{n,w}(E_2 \cup E_{\text{init}}). \quad (51)$$

If $e \in E_{\text{init}}$, both sides of (51) become zero. Hence we can safely assume that $e \notin E_{\text{init}}$. To shorten our notation let us define $E_i^* \triangleq E_i \cup E_{\text{init}}$ for $i = 1, 2$. Therefore (51) can be rewritten as,

$$\log \text{tree}_{n,w}(E_1^* \cup \{e\}) - \log \text{tree}_{n,w}(E_1^*) \geq \log \text{tree}_{n,w}(E_2^* \cup \{e\}) - \log \text{tree}_{n,w}(E_2^*). \quad (52)$$

Recall that by assumption (V, E_{init}) is connected. Thus (V, E_i^*) is connected for $i = 1, 2$, and we can apply Lemma 3.1 on both sides of (52). After doing so we have to show that

$$\log(1 + w_e \Delta_e^{G_1}) \geq \log(1 + w_e \Delta_e^{G_2}) \quad (53)$$

where $G_i \triangleq (V, E_i \cup E_{\text{init}}, w)$ for $i = 1, 2$. It is easy to see that (53) holds iff $\Delta_e^{G_1} \geq \Delta_e^{G_2}$. Now note that

$$\Delta_e^{G_1} - \Delta_e^{G_2} = \mathbf{a}_e^\top (\mathbf{L}_{G_1}^{-1} - \mathbf{L}_{G_2}^{-1}) \mathbf{a}_e \geq 0 \quad (54)$$

since $\mathbf{L}_{G_2} \succeq \mathbf{L}_{G_1}$ (G_1 is a spanning subgraph of G_2), and therefore according to Lemma A.1 $\mathbf{L}_{G_1}^{-1} \succeq \mathbf{L}_{G_2}^{-1}$. \square

Proof of Theorem 4.3. First note that (23) directly follows from Theorem 3.1 since $\mathbf{L}(\boldsymbol{\pi})$ is the reduced weighted Laplacian matrix of G_\bullet after scaling its edge weights by the sampling probabilities π_1, \dots, π_s . To prove (24) consider the following indicator function,

$$\mathbb{1}_{E(\mathcal{H})}(e) = \begin{cases} 1 & e \in E(\mathcal{H}), \\ 0 & e \notin E(\mathcal{H}). \end{cases} \quad (55)$$

Now note that $\mathbb{1}_{E(\mathcal{H})}(e_i) \sim \text{Bern}(\pi_i)$ for $i = 1, \dots, s$. Therefore,

$$\mathbb{E}_{\mathcal{H} \sim \mathbb{G}(G_\bullet, \pi_\bullet)}[|E(\mathcal{H})|] = \mathbb{E}_{\mathcal{H} \sim \mathbb{G}(G_\bullet, \pi_\bullet)}\left[\sum_{i=1}^s \mathbb{1}_{E(\mathcal{H})}(e_i)\right] \quad (56)$$

$$= \sum_{i=1}^s \mathbb{E}_{\mathcal{H} \sim \mathbb{G}(G_\bullet, \pi_\bullet)}\left[\mathbb{1}_{E(\mathcal{H})}(e_i)\right] \quad (57)$$

$$= \sum_{i=1}^s \pi_i \quad (58)$$

$$= \sum_{e_i \in \mathcal{M}^+} \pi_i + \sum_{e_j \in E_{\text{init}}} 1 \quad (59)$$

$$= \|\pi\|_1 + |E_{\text{init}}|. \quad (60)$$

□

Proof of Theorem 4.4. This theorem is a direct application of Chernoff bounds for Poisson trials of independently sampling edges from \mathcal{M}^+ with probabilities specified by π^* . □

Generalizing Theorems 3.1 and 4.3

The following theorem generalizes Theorem 3.1 (and, consequently, Theorem 4.3). Theorem A.1 provides a similar interpretation for the convex relaxation approach designed by Joshi and Boyd [14] for the sensor selection problem with linear measurement models.

Theorem A.1. Let $\{(\mathbf{y}_i, \mathbf{z}_i)\}_{i=1}^m$ be a collection of m pairs of vectors in \mathbb{R}^n such that $m \geq n$. Furthermore, let s_1, \dots, s_m be a collection of m independent random variables such that $s_i \sim \text{Bern}(p_i)$ for some $p_i \in [0, 1]$. Then we have,

$$\mathbb{E}_{\substack{s_i \sim \text{Bern}(p_i) \\ \forall i \in [m]}} \left[\det \left(\sum_{i=1}^m s_i \mathbf{y}_i \mathbf{z}_i^\top \right) \right] = \det \left(\sum_{i=1}^m p_i \mathbf{y}_i \mathbf{z}_i^\top \right). \quad (61)$$

Proof. Let $\mathcal{S}_n \triangleq \binom{[m]}{n}$ be the set of all n -subsets of $[m]$. According to the Cauchy-Binet (C-B) formula we have

$$\mathbb{E}_{\substack{s_i \sim \text{Bern}(p_i) \\ \forall i \in [m]}} \left[\det \left(\sum_{i=1}^m s_i \mathbf{y}_i \mathbf{z}_i^\top \right) \right] \stackrel{\text{C-B}}{=} \mathbb{E}_{\substack{s_i \sim \text{Bern}(p_i) \\ \forall i \in [m]}} \left[\sum_{\mathcal{Q} \in \mathcal{S}_n} \det \left(\sum_{i \in \mathcal{Q}} s_i \mathbf{y}_i \mathbf{z}_i^\top \right) \right] \quad (62)$$

$$= \sum_{\mathcal{Q} \in \mathcal{S}_n} \mathbb{E}_{\substack{s_i \sim \text{Bern}(p_i) \\ \forall i \in [m]}} \left[\det \left(\sum_{i \in \mathcal{Q}} s_i \mathbf{y}_i \mathbf{z}_i^\top \right) \right]. \quad (63)$$

Now note that $|\mathcal{Q}| = n$ and $\text{rank}(\mathbf{y}_i \mathbf{z}_i^\top) = 1$. Therefore $\det \left(\sum_{i \in \mathcal{Q}} s_i \mathbf{y}_i \mathbf{z}_i^\top \right)$ is non-zero iff $s_i = 1$ for all $i \in \mathcal{Q}$. Thus for every $\mathcal{Q} \in \mathcal{S}_n$,

$$\det \left(\sum_{i \in \mathcal{Q}} s_i \mathbf{y}_i \mathbf{z}_i^\top \right) = \begin{cases} d_{\mathcal{Q}} \triangleq \det \left(\sum_{i \in \mathcal{Q}} \mathbf{y}_i \mathbf{z}_i^\top \right) & \text{with probability } p_{\mathcal{Q}} \triangleq \prod_{i \in \mathcal{Q}} p_i, \\ 0 & \text{with probability } 1 - p_{\mathcal{Q}}. \end{cases} \quad (64)$$

Taking the expectation yields

$$\mathbb{E}_{\substack{s_i \sim \text{Bern}(p_i) \\ \forall i \in [m]}} \left[\det \left(\sum_{i \in \mathcal{Q}} s_i \mathbf{y}_i \mathbf{z}_i^\top \right) \right] = p_{\mathcal{Q}} d_{\mathcal{Q}}. \quad (65)$$

Replacing (65) in (63) results in

$$\mathbb{E}_{\substack{s_i \sim \text{Bern}(p_i) \\ \forall i \in [m]}} \left[\det \left(\sum_{i=1}^m s_i \mathbf{y}_i \mathbf{z}_i^\top \right) \right] = \sum_{\mathcal{Q} \in \mathcal{S}_n} p_{\mathcal{Q}} d_{\mathcal{Q}} = \sum_{\mathcal{Q} \in \mathcal{S}} \det \left(\sum_{i \in \mathcal{Q}} p_i \mathbf{y}_i \mathbf{z}_i^\top \right). \quad (66)$$

Noting that the RHS in (66) is the Cauchy-Binet expansion of $\det \left(\sum_{i=1}^m p_i \mathbf{y}_i \mathbf{z}_i^\top \right)$ concludes the proof. \square